Oldest-Old Mortality Rates and the Gompertz Law: A Theoretical and Empirical Study Based on Four Countries[*]

高高齡老人死亡率與 Gompertz 模型：理論與實證研究

Jack C. Yue[**]

余清祥

ABSTRACT

Testing the Gompertz law (i.e. law of geometrical progression) for elderly mortality rates has been discussed in the literature for a long time, but tests based on a set of yearly age-specific data have not been fully explored yet. In the first part of this paper, we propose a standard operating procedure for testing the Gompertz assumption using yearly age-specific mortality data. Methods used in the procedure include estimation of parameters in the Gompertz law and their standard errors via bootstrap simulation. In addition to the oldest-old (i.e. ages 80 and above) data in Japan, Sweden, France, and the U.S. (Data sources: *Berkeley Mortality Database* and Kannisto, 1994), a simulation study is used to demonstrate the validity of the proposed procedure. However, in practice the period of data collection is often prolonged to 5 or 10 years in order to accumulate sufficient sample sizes. However, a longer data collection period is likely to mix data with different attributes and cause problems in the parameter estimation. Thus, in the second part of this paper, we discuss the impacts of the data collection period and population sizes on the testing results.

Key Words: Gompertz's Law; Mortality Rates of the Elderly; Mortality Projection; Bootstrap; Simulation; Maximum Likelihood; Weighted Least Square; Nonlinear Maximization; Graduation

---

摘　要

　　高齡老年人(65歲以上)的死亡率歷來都被認為滿足 Gompertz 模型，實證中有不少高齡資料以此假設來配適，但因受限於資料量的不足，鮮少有研究以統計檢定的觀點檢視 Gompertz 模型是否合適，因而至今仍有正反兩派意見支持與反對將老年人死亡率以 Gompertz 模型來配適或推估。有鑑於此，本研究提出一套檢測 Gompertz 假設的方法，包括模型的檢定、參數的估計、及藉由電腦模擬(拔靴法)得到參數估計值的標準差。值得一提的是本方法也可用於僅有單一年度各年齡別的死亡資料，有別於過去研究中統計檢定需要兩組以上(也就是兩個以上的年度)的各年齡別的死亡資料。本文除了實證部分採用日本、瑞典、法國、美國這四個國家80歲以上的死亡資料(資料來源：美國柏克萊大學死亡率資料庫與 Kannisto, 1994)，也藉由電腦模擬檢驗本文提出的方法是否正確。除此之外，實證中常因樣本數的不足，而將原先單一年度的資料延長成合併數個年度(通常為五年或十年)的資料，本文第二部分也將探討這種將不同特性的樣本合併對檢測 Gompertz 假設可能造成的影響。

關鍵字：　Gompertz 模型、高齡死亡率、死亡率的推估、拔靴法、電腦模擬、最大概似估計、加權最小平方法、非線性估計、修勻

1. **INTRODUCTION**

British actuary Benjamin Gompertz proposed a simple formula in 1825, for describing the mortality rates of the elderly. He observed a law of geometrical progression (i.e. exponential rise) in death rates based on 19th century people in England, Sweden, and France between ages 20 and 60.

In terms of actuarial notation, this formula can be expressed as

$$\mu_x = BC^x, \tag{1}$$

where $B > 0$ and $C > 1$ are model parameters, $x > 0$ is the age, and μ_x is the force of mortality at age x. Under the Gompertz assumption, the (conditional) probability that an individual now aged x would survive to age $x+1$, denoted by p_x, is

$$p_x = e^{-\int_x^{x+1} \mu_t dt} = e^{-\int_x^{x+1} BC^t dt}$$

$$= e^{-BC^x(C-1)/\log C}, \tag{2}$$

or equivalently, $\log p_x = -BC^x(C-1)/\log C$ and

$$\log p_{x+1}/\log p_x = C. \tag{3}$$

Usually, because of its simplicity, the ratio of $\log p_x$ in (3) is used more often than p_x in (2) to check if the Gompertz assumption is valid.

If the mortality rates of the elderly follow the Gompertz law (i.e. a parametric assumption) and the estimation of the parameters (B and C) is not too complicated, then the mortality projection would be more straightforward than using regular non-parametric methods. This is one of the main reasons why a parametric approach has been used overwhelmingly in mortality projections, especially for elderly mortality rates (for discussions of mortality projection, see for

example, Pollard, 1987).

The focus of applying the Gompertz law has been on modeling elderly mortality rates, but its application is not restricted to mortality rates. For example, Finch and Pike (1996) examined the maximum life span prediction by assuming that the age-related mortality of an adult population follows the Gompertz law. They found that the population size has little influence on the maximum life span, while the model parameter C has much more influence. The Gompertz model can also be used to model fertility rates. Booth (1984) used the Gompertz model to estimate the pattern of fertility rates. Based on the simulation and Swedish data, he found that the Gompertz model produces very good estimates when the fertility rate is declining.

Although the Gompertz law has received much attention, there are still no standard methods to test if a set of yearly age-specific mortality data satisfies this law. (Prentice and El Shaarawi, 1973, derived a test of fit for the Gompertz force of mortality based on having several observed death rates at each age class.) The lack of testing procedures for the Gompertz law is generally due to the data size and data quality, in particular, for the elderly. First, because usually only a small number of individuals survive to advanced ages (i.e., a small sample size), this would cause non-negligible random fluctuations in the estimation. Thus, in order to collect enough death data, the period of data collection is often prolonged to 5 or 10 years (for example, Kannisto, 1994), unlike in the case of regular life tables where usually 3 years of death data are involved. However, the mortality tends to improve over time and the improvement of elderly mortality rates is especially significant (see for example, Willets, 1999). A 10-year period of data collection is likely to mix together sets of data with different attributes, i.e. the data are not homogeneous, and possibly cause problems in parameter estimation and mortality projection.

The second problem (data quality) usually is the inaccuracy in the reporting of age. In

general, it is not easy to trace and verify the age of an elderly person, especially a very old individual. For example, Wilmoth et al. (1996) discussed the case of verifying a 114-year-old man. The inaccuracy in the reporting of age usually comes from an overstatement and heaping on ages that are multiples of five or ten. Overstatement of ages is generally difficult to identify, while the age heaping can be detected and adjusted by methods such as Meyer's method with certain assumption (see for example, Brown, 1991).

In this paper we will focus on the testing procedure of the Gompertz law, if there is only a set of yearly age-specific data available. When dealing with empirical studies, such as checking whether the mortality rates of certain countries follow the Gompertz law, we assume that there are no problems in data quality. The issue of data quality is beyond the scope of the current study. Therefore, it should be noted that the results of empirical studies are based on the assumption of good data quality, and it should not be overlooked. Also, in order to avoid the problem of in-homogeneous data (due to a long data collection period) and insufficient sample sizes, the data used in this study are yearly based and have moderate sample sizes. For the following sections, we will first state the assumption and methodology of this study (Section 2), and then introduce the idea of our proposed testing procedure and its merit (Section 3).

2. METHODLOGY

In this paper, given that the Gompertz law of mortality is satisfied, we assume that for people now aged x, the number of deaths (denoted by D_x, a random variable) before reaching age $x+1$ follows a binomial distribution, $B(n_x, q_x)$, where $q_x = 1 - p_x$, p_x is given in (2),

and n_x is the number of people now aged x. We also assume that D_x and D_y are independent for $x \neq y$. Let d_x denote the observed number of D_x. If n_x is sufficiently large, then by the Central Limit Theorem,

$$\frac{d_x - n_x q_x}{\sqrt{n_x q_x (1 - q_x)}} \approx N(0,1),$$

or equivalently, letting $\hat{p}_x = 1 - d_x / n_x$,

$$\frac{\hat{p}_x - p_x}{\sqrt{p_x (1 - p_x) / n_x}} \approx N(0,1),$$

and

$$\sum_{x=a}^{b} \frac{(\hat{p}_x - p_x)^2}{p_x (1 - p_x) / n_x} \approx \chi^2 (b - a + 1), \tag{4}$$

where a and b are the lowest and highest ages of the mortality data. Therefore, if the mortality data follow the Gompertz law, then we can use chi-square goodness-of-fit to test the overall significance. This solves the problem of testing once we have the estimated values.

We now need to find a reliable method of parameter estimation in the Gompertz law (i.e. B and C). From (3), the ratio of $\log p_{x+1}$ to $\log p_x$ is a constant, meaning that $\log(-\log p_x)$ is a linear function of x, or

$$\log(-\log p_x) = \log B + \log(C-1) - \log(\log C) + x \log C = \alpha + \beta x, \tag{5}$$

where $\alpha = \log B + \log(C-1) - \log(\log C)$ and $\beta = \log C$. Therefore, the Ordinary Least Squares (OLS) can be used to solve for α and β. Note that because exposures of the elderly differ a lot for different ages, a *Weighted Least Squares* (WLS) procedure is recommended, i.e.

$$\min_{\alpha, \beta} \sum_x w_x \left(\log(-\log p_x) - \alpha - \beta x \right)^2, \tag{6}$$

where w_x is the weight.

There are limitations in using WLS and it cannot be used when the observed values of p_x are either 0 or 1. A more general setting of WLS is *Nonlinear Maximization* (NM). Since (2) gives the parametric form of p_x, similar to WLS, the estimates of B and C can be solved via minimizing

$$\min_{B,C} \sum_x w_x \left(p_x - e^{-BC^x(C-1)/\log C} \right)^2 , \qquad (7)$$

where w_x is the weight. The NM can be solved via numerical methods (such as the Newton-Raphson iteration and Simplex algorithm, to name just a few), and in this study, we use the function *nlminb* (a local minimizer for smooth nonlinear functions subject to bound-constrained parameters) in S-Plus to solve for the parameters B and C.

Another choice of parameter estimation is the *Maximum Likelihood Estimation* (MLE). Since we assume that the number of deaths for each age follows a binomial distribution, the logarithm of the joint likelihood function can be written as

$$\log L(B,C) = \sum_x (n_x - d_x) \log p_x + d_x \log(1 - p_x) . \qquad (8)$$

Following the same idea in NM for finding the minimum and plugging into (2) for p_x, the MLE is the one satisfying

$$\min_{B,C} \sum_x \left[(n_x - d_x) BC^x(C-1)/\log C - d_x \log\left(1 - e^{-BC^x(C-1)/\log C} \right) \right]. \qquad (9)$$

The MLE in this study is also solved via *nlminb* in S-Plus, similar to NM.

Note that, unlike the WLS, NM and MLE can deal with cases where the observed values of p_x are 0 or 1. In this study, when the WLS is used and there are 0's or 1's in the observed values of p_x, which implies a small number of observations (i.e. usually for older age groups), then

this age group is omitted from the estimation. On the other hand, because the minimums in (7) and (9) are achieved by numerical methods and iterations, it is possible that the numerical solutions do not converge to the correct values. To avoid incorrect convergence, the initial values of parameters B and C in using *nlminb* need to be carefully chosen. Also, in order to judge which method gives the best estimation, a simulation will be used to evaluate the performances in the following section.

3. RESULTS

In this section we first propose a procedure for checking the Gompertz law. Our proposed procedure consists of the following three steps:

1. *Checking if the data violate the Gompertz law. If the data violate the assumption,*
    stop; otherwise, go to the next step.
2. *Estimation of the model parameters.*
3. *Goodness-of-fit test.*

We will explain how and why the proposed procedure works, step by step.

Step 1. Checking the Gompertz assumption

If the data follow the Gompertz law, the ratios of $\log p_{x+1}$ to $\log p_x$ shall be fairly close to the parameter C.   Therefore, the confidence intervals of C constructed via the bootstrapping simulation at each age and their intersection shall contain the true C with a probability related to the confidence coefficients. However, because the confidence intervals constructed in different ages are not independent, it is very difficult to determine the confidence

coefficient of the intersection of the bootstrapping confidence intervals.   Instead, we use 100%

confidence intervals.   Therefore, if the intersection of the bootstrap confidence intervals is

empty, then the Gompertz assumption is violated; otherwise, we "**do not reject**" (instead of

"**accept**") the assumption that the data follow the Gompertz law.   It shall be noted that the

proposed method above is extremely conservative and shall be treated as a preliminary check.

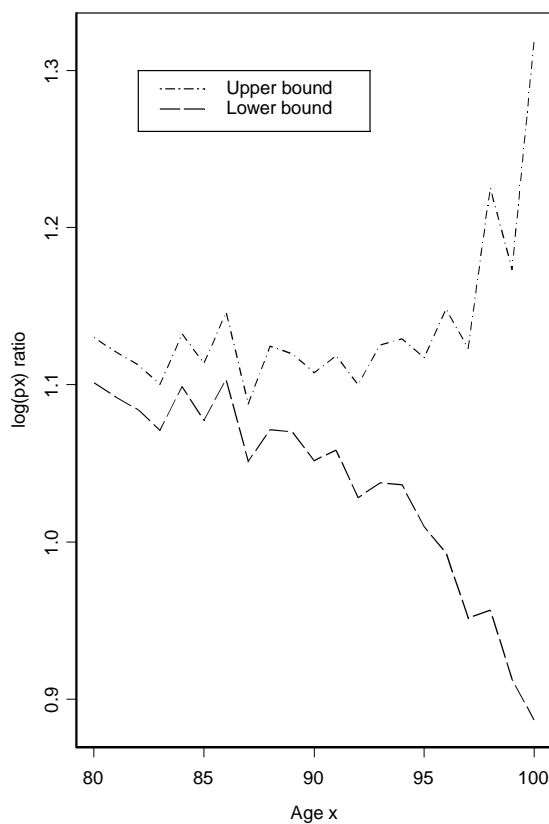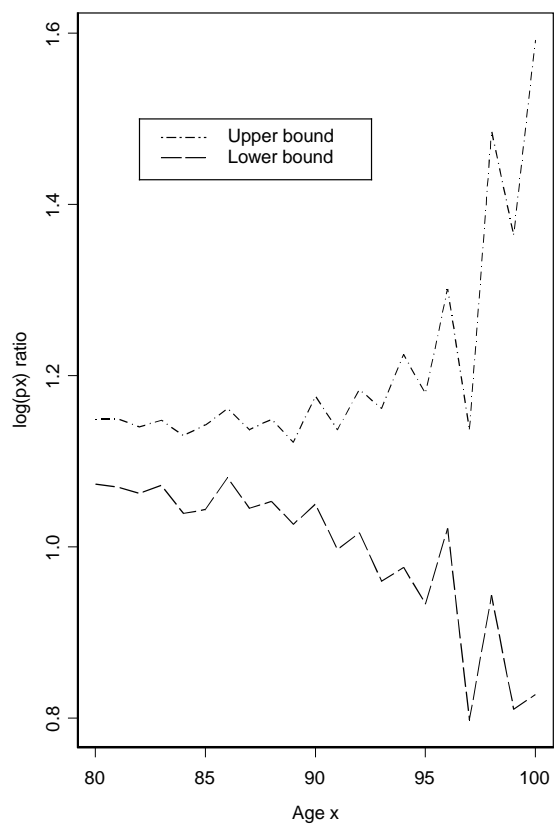Fig 1-1. Bootstrap C.I. for Japanese Male   Fig 1-2. Bootstrap C.I. for Swedish Male

We use Japanese and Swedish elderly data (both sets are age-specific data, collected in the

time period: 1980-1990) in Kannisto (1994) as a demonstration. The bootstrapping simulations

were conducted based on generating random numbers from binomial distribution $B(n_x, \hat{q}_x)$ for

ages $x \geq 80$, where n_x and $\hat{q}_x$ are the observed number of population size and the observed

mortality rate for age x in 1980-1990 for Japan and Sweden. Figures 1-1 and 1-2 are the

bootstrapping confidence intervals for the Japanese male and Swedish male, with 1,000 bootstrapping replications. The intersection of bootstrapping confidence intervals is empty in the case of the Japanese male, which indicates that the mortality rates of the Japanese male violate the assumption of the Gompertz law. On the other hand, in the case of the Swedish male, the intersection apparently is empty and we do not reject the assumption that the data of the Swedish male follow the Gompertz law.   Therefore, we suggest that only the parameter estimation of Sweden's male cases will be continued in the next step.

Note that although bootstrapping confidence intervals are used to determine if the mortality data follow the Gompertz law, we do not recommend using this method to determine the parameter C.   This is due to the reason that we cannot decide upon the confidence coefficient, and it is also because the intersection of confidence intervals contains more than one point, as seen in the case of the Swedish male.   Since the choice of C is not unique, the selection of a possible C would likely become subjective.

Table 1. Model parameters of Gompertz's law for Japan and Sweden

	Japan		Sweden	
	Male	Female	Male	Female
B	.0000725	.0000110	.0000799	.0000145
C	1.093146	1.113206	1.092694	1.109986
Ave. n_x	282335	493072	44996	75665

We also use the simulation to check the power of the preliminary test for the model assumption.   In other words, we want to know the probability of false rejection (i.e. type-I error) for applying the bootstrapping simulation in Step 1, provided that the Gompertz law is true.   It should be noted that the assumption used here should not be mixed with the empirically testing

results in Step 1 (i.e. Figures 1-1 and 1-2). Using the populations records in Japan and Sweden (for both male and female) from Kannisto (1994) and the values of B and C given in Table 1, 1,000 simulation runs are conducted to see if the data conform to the Gompertz law, and how many times we would observe empty intersection (like in Figure 1-1). Note that the values of B and C in Table 1 are the WLS estimators of Japanese and Swedish death records in 1980-1990 from Kannisto. Among 1,000 simulation runs for all four cases, we observed no runs with an empty intersection. This implies that the preliminary test for the Gompertz law is very reliable (i.e. type-I error equals 0) and thus we would expect a non-empty intersection if the data follow the Gompertz law. In other words, if the intersection of bootstrapping confidence intervals for a data set is empty, then it is very likely that this data set does not follow the Gompertz law.

There are two reasons why we propose a pre-test (Step 1) on the Gompertz law. First, if the mortality rates apparently fail the assumption, then we only need to proceed with the pre-test and see whether the Gompertz law is satisfied. This can save us time and effort for not doing the parameter estimation in the next step. Based on the simulation result of zero type-I errors, we know that there is little risk for falsely rejecting the Gompertz assumption and it is safe to proceed with the pre-test. Secondly, the intersection of bootstrapping confidence intervals can provide a range of possible values for C, and these values can be used to double check with the estimate derived from Step 2.

Step 2. Parameter Estimation.

As mentioned in the previous section, we can use WLS, NM, and MLE to estimate the model parameters. In this section, we conduct a simulation, based on the mortality data from Japan and Sweden, to evaluate the performance of these three methods. The model parameters

and population sizes of each case are shown in Table 1.

Using either the WLS or NM involves the choice of weights (w_x). For WLS, $w_x = 1$, $\sqrt{n_x}$, n_x, and $\log n_x$ are all selected, while $w_x = 1$, $\sqrt{n_x}$, and n_x are selected for NM. (The reason that we choose $w_x = \log n_x$ is due to the fact that $\log p_x$ is used in WLS.) Here, in order to evaluate the performance of these estimators, we consider the following loss function:

$$L(\hat{\theta},\theta) = Var(\hat{\theta}) + E(\hat{\theta} - \theta)^2,$$

which includes both the variance and $(\text{bias})^2$ for the estimator $\hat{\theta}$, where $\theta = B$ or C.

Tables 2-1 and 2-2 list the simulation results of 1,000 replications for the four cases in Table 1. Since the population sizes in Japan are larger, the simulated losses are thus smaller, as expected. Except for MLE, the case $w_x = n_x$ performs the best (with respect to loss), for both WLS and NM. The loss increases as the weight n_x decreases. In general, if WLS or NM is used, we would suggest readers to use $w_x = n_x$, because it can reflect the sample sizes in each age group and the estimation will not be influenced by a small number of individuals at higher ages.

As for the estimation method, MLE is the best among all methods, while NM is slightly better than WLS though the edge is very small. Our result is similar to that in Thatcher (1990), where he found that MLE has the closest fit for parameters in the Heligman and Pollard model. Note that because the standard errors of the estimates for B and C are fairly small, the majority of simulated losses comes from the term $(\text{bias})^2$. In addition, the standard errors of the estimates for B and C can also be derived from bootstrapping simulation, similar to the idea in Step 1, and they are thus omitted.

Table 2-1. Simulated loss of parameter B for WLS, NM, and MLE (unit: 1)

Method: Weights		Japan		Sweden	
		Male	Female	Male	Female
WLS:	n_x	.000485	.000309	.002884	.001741
	$\sqrt{n_x}$	.000928	.000455	.003998	.002090
	$\log(n_x)$	.029293	.012961	.034652	.020854
	1	.207920	.051505	.126977	.076628
NM:	n_x	.000459	.000295	.002570	.001711
	$\sqrt{n_x}$	.005189	.003433	.015640	.012701
	1	1.36493	.373927	.652265	.680022
MLE	1	.000418	.000248	.002498	.001419

Table 2-2. Simulated loss of parameter C for WLS, NM, and MLE (unit: 0.001)

Method: Weights		Japan		Sweden	
		Male	Female	Male	Female
WLS:	n_x	.000146	.000093	.000859	.000515
	$\sqrt{n_x}$	.000280	.000138	.001168	.000616
	$\log(n_x)$	.007751	.003696	.009002	.005687
	1	.039220	.013698	.028503	.018376
NM:	n_x	.000138	.000088	.000759	.000497
	$\sqrt{n_x}$	.001465	.000950	.004239	.003402
	1	.127820	.005851	.099599	.090268
MLE	1	.000126	.000074	.000729	.000416

Step 3. Goodness-of-fit test

In addition to the loss function, we also conduct the chi-square goodness-of-fit test, as stated in (4), to evaluate the performance of WLS, NM, and MLE. Table 3-1 shows the numbers of rejections among 1,000 simulation replications for the four cases in Table 1. The

results in Table 3-1 are very similar to those in Tables 2-1 and 2-2, and only four cases have zero errors: WLS with $w_x = n_x$ and $w_x = \sqrt{n_x}$; NM with $w_x = n_x$; and MLE. These four cases also have the smallest losses in Tables 2-1 and 2-2. From the view of hypothesis testing, the numbers in Table 3-1 can also be interpreted as a type-I error, if the null hypothesis is whether the mortality data satisfy the Gompertz assumption. Following the usual setting for the type-I error, i.e. 0.05, only the four cases with zero errors can be used to check the Gompertz assumption.

Table 3-1. Number of rejections among 1,000 simulations for WLS, NM, and MLE

Method: Weights		Japan		Sweden	
		Male	Female	Male	Female
WLS:	n_x	0	0	0	0
	$\sqrt{n_x}$	0	0	0	0
	$\log(n_x)$	480	371	114	76
	1	790	689	392	349
NM:	n_x	0	0	0	0
	$\sqrt{n_x}$	159	145	46	51
	1	943	934	836	831
MLE	1	0	0	0	0

In addition to the type-I error, we can also use a simulation to evaluate the type-II error. In particular, we assume that the null hypothesis is $H_0 : B = B_0 \ \& \ C = C_0$, where the values of B_0 and C_0 are given in Table 1. Assume that the true parameters are not (B_0, C_0) and we are interested in knowing the probability of not rejecting H_0 (i.e. type-II error). The true parameters considered are $((1 - p\%)B_0, C_0)$, where $p = 1$, 0.5, 0.2, 0.1, and $(B_0, (1 - q\%)C_0)$, where $q = 0.1$, 0.01, 0.005, 0.001. The reason why the choices of q are smaller is because

the proposed test is more sensitive to the change in C. (This can be seen from (2) as well.)

Table 3-2 lists the simulation results based on 1,000 replications and significance level (i.e. type-I error) of 0.05. Since the type-II error is the probability of false acceptance, the smaller the numbers in Table 1 are, the larger the type-II errors will be. Note that the Japanese populations are larger than those of Sweden, and thus the type-II errors are generally larger in the Swedish cases. Similarly, the male populations also have larger type-II errors because of smaller sizes. Also, because the parameter C is the main attribute of the Gompertz law (i.e. $\mu_x = BC^x$, exponential rise in C), the type-II errors of C with respect to the difference in percentage are generally more sensitive.

Table 3-2. Number of rejections among 1,000 simulations when the null hypothesis is wrong

Parameter: Difference		Japan		Sweden	
		Male	Female	Male	Female
B:	1%	1,000	1,000	15	95
	0.5%	188	485	0	0
	0.2%	0	0	0	0
	0.1%	0	0	0	0
C:	0.1%	1,000	1,000	1,000	1,000
	0.01%	1,000	1,000	5	25
	0.005%	46	198	0	0
	0.001%	0	0	0	0

4. DISCUSSION

In this section we apply our proposed procedure on the mortality data in the Berkeley Mortality Database and check if the oldest-old mortality rates satisfy the Gompertz law. This database was established in 1997 by Professor John R. Wilmoth of the Department of

Demography at the University of California, Berkeley and it contains life tables and demographic records for national populations in the following four countries: France, Japan, Sweden, and the United States. As mentioned in first section, these four countries have fairly large yearly populations; see Table 4.1 for the populations of age 65 and over in 1993 as a demonstration. This paper chooses all four countries from this database to check for the proposed procedure (numbers in the parenthesis are the data periods): France (1971-1995), Japan (1971-1996), Sweden (1971-1999), and the U.S. (1971-1992).

It should be noted that in order to acquire coherent estimates of mortality rates, in all four countries the mortality rates are calculated via dividing the numbers of deaths by the numbers of population estimates at each age and gender. Because we are interested in the oldest-old mortality, the starting age is set to 80 in all four countries (it makes no difference if the starting age is chosen as 70, 60, or 50 year-olds). The highest age included in the analysis varies from 102 to 109, depending on the data availability.

Table 4.1. Population of ages 65+ in 1993 for France, Japan, Sweden, U.S., and Taiwan

	France	Japan	Sweden	U.S.	Taiwan	Taipei
Male	331.9	662.1	65.1	1318.5	54.9	8.7
Female	503.9	968.8	88.4	1937.6	63.8	11.2
Total	835.8	1630.9	153.5	3256.0	118.7	19.9

(Unit: 10,000)

The testing results for these four countries are as follows: The mortality data from France, Japan, and the U.S. all fail the assumption of the Gompertz law, according to the testing procedures proposed in Section 3. Some of the mortality data in Sweden, mostly in the 1970s (and some in the late 1980s and 1990s) and for the females, do in fact satisfy the assumption of

the Gompertz law.   Note that the testing results are the same if the highest age included in the analysis is reduced to 95 or 100.   Based on these findings, we think that directly applying the Gompertz law of mortality to the oldest-old (ages 80 and over) group is questionable.   Still, we cannot conclude that the Gompertz law is an inappropriate assumption.   In the following, we shall further investigate the possible causes of why the Gompertz law fails in our empirical studies.

Even if the mortality profiles follow the Gompertz assumption, it is still possible to fail the proposed test.   One of the possible causes is the difference of mortality profiles between urban and rural areas.   If the mortality rates of urban and rural areas are not similar, then it would be inappropriate to mix them together and use a single Gompertz function to model them. For example, Table 4.2 shows the life expectancy in the areas of Taiwan.   Obviously, the life expectancy of urban people in Taipei city (for both male and female) is much higher than those of the 16 counties, and the difference has been increasing in recent years, especially for the male. However, because the population sizes in Taiwan are much smaller than those in the U.S., Japan, and France (Table 4.1), we do not reject the Gompertz assumption if we combine the mortality data of Taipei city and the 16 counties. However, if we fix the mortality rates and increase the population sizes for each age in Taiwan by 30 times, then the Gompertz assumption is indeed rejected.   This could be one of the reasons that why only some of the mortality rates in Sweden (and Taiwan) satisfy the Gompertz assumption, while those of France, Japan, and the U.S. do not. Note that the population sizes required to reject the Gompertz assumption depend on the mortality profiles and the population structures as well.

The discrepancies of mortality profiles for a country are not restricted to living areas (between urban and rural areas). Other factors, such as races/ethnicity (e.g. black and white

populations in the U.S.), lifestyles (e.g. smoking habits), and income levels also have large impacts on mortality rates. When the population sizes of these sub-populations (classified according to these factors) increase, the Gompertz law becomes an unrealistic assumption for the mortality rates. Therefore, if the sub-population sizes are big and obviously the subpopulations have very different mortality profiles (i.e. we have heterogeneous sub-populations), then we suggest checking the Gompertz assumption on each sub-population, instead of the whole population.

Table 4.2. Life Expectancy in Taiwan (Unit: Year)

	16 Counties Male	16 Counties Female	Taipei Male	Taipei Female
1991	71.20	76.70	75.88	80.22
1992	71.19	76.76	75.95	80.54
1993	71.16	77.14	75.99	80.83
1994	71.51	77.46	76.18	80.95
1995	71.47	77.44	76.18	81.08
1996	71.48	77.42	76.37	81.14
1997	71.16	77.63	76.51	80.96
1998	71.45	77.84	76.56	81.20
1999	71.52	77.81	76.84	81.55
2000	71.70	78.09	76.97	81.62

5.   CONCLUSION

In this paper we introduce a testing (and also a standard operating) procedure for checking whether a set of yearly age-specific data follows the Gompertz law. Via simulation and the goodness-of-fit test, we have shown that this procedure is very reliable in verifying the Gompertz assumption. (In other words, if the null hypothesis is whether the data set follows the Gompertz

law, then the proposed procedure has a very small type-I error.)   If (based on the preliminary check) we do not reject that a data set follows the Gompertz law, then we suggest using MLE, NM with $w_x = n_x$, or WLS with $w_x = n_x$ in order to estimate the model parameters B and C, since these three estimates have the smallest losses. In addition, the chi-square goodness-of-fit test also confirms that these estimates are very reliable (if based on type-I error).   However, since MLE and NM are used less often, we suggest using WLS with $w_x = n_x$ in practice.

It shall be noted that, although the proposed procedure is a reliable tool for testing the Gompertz law, we do not recommend readers look at the testing results only.   As pointed out in the first section, two questions ("*Are the data reliable?*" and "*Are the data homogeneous?*") need to be answered before the proposed procedure is applied.   The first question is particularly difficult to answer and verify, since usually there is little or no information available about how the mortality data were collected.

The second question of whether the data are homogeneous is even more interesting and needs a detailed discussion. Note that small population sizes would create large fluctuations in an estimation and likely lower the testing power (i.e. $1 - \beta$, where β is the type-II error) of the proposed procedure.   For example, for the mortality data of a Taiwanese male in 1993, 77 times out of 100 replications from the confidence intervals from 1,000 bootstrapping simulations fail to satisfy the Gompertz law.   If the observed mortality rates are fixed and the population sizes at each age are reduced to 1/5 of the original size, then none of 100 replications fail the test.   This (and the results in Table 3.2) could be used to explain why most past studies do not show obvious evidence to reject the Gompertz assumption (since the population sizes are not too big).   Thus, in order to increase the power of testing, the yearly data of the four countries chosen in this paper's empirical studies all have large population sizes.

There is a trade-off in choosing data sets with large populations, and similar to a data collecting period of 5 or 10 years, it is likely one will mix two (or more than two) different populations together. Since the goal is to accumulate a large yearly population of elderly, the overall population and their living areas would be large as well. This indicates that discrepancies might exist within the large population. We suggest conducting an explanatory data analysis on the whole population and make sure the homogeneity of mortality profiles exists, before applying the proposed procedure for testing the Gompertz assumption.

There are alternatives for dealing with the problem of large random fluctuations due to small population sizes. For example, before going through the testing procedure, we can use graduation (data smoothing) methods to reduce the fluctuations. Another alternative is to pool the raw data into 5-year (or 10-year) age groups, like Wilmoth (1995) did in fitting the data to Coale and Kisker's model. These two methods certainly are not perfect and have drawbacks. The graduation methods might alter the original data attributes (see London, 1985, or Yue, 1997, for a detailed discussion of graduation methods).

The pooling method has its problems as well, since the Gompertz law is for a single age (i.e., the ratio $\log p_x$ in equation (3)) and we still need to break down the death data from 5-year groups into single age groups. Of course, we can use cumulative mortality rates of a 5-year age group, instead of mortality rates of a single age, to check this assumption, and methods like chi-square goodness-of-fit can be used. However, this kind of method would require more complicated computations, and the groupings (such as ages 80-84 or 81-85) might influence the testing result and make the testing result more controversial (like in the usual chi-square tests).

BIBLIOGRAPHY

Brass, W., Perspectives in Population Prediction: Illustrated by the Statistics of England and Wales, *Journal of Royal Statistical Society*, 1974, Series A, 137, 532-570.

Carnes, B.A. , Olshansky, S.J., and Grahn, D., Continuing the Search for a Law of Mortality, *Population and Development Review*, 1996, 22, No. 2, 231-264.

Coale, A. and Guo, G., Revised Regional Model Life Tables at Very Low Levels of Mortality, *Population Index*, 1989, 55(4), 613-643.

Finch, C.E. and Pike, M.C., Maximum Life Span Predictions From the Gompertz Mortality Model, *Journal of Gerontology: Biological Sciences*, 1996, 51A, No. 3, 183-194.

Garg, M.L., Rao, B.R., and Redmond, C.K., Maximum-Likelihood Estimation of the Parameters of the Gompertz Survival Function, *Applied Statistics*, 1970, 19(2), 152-159.

Heligman, L. and Pollard, J.H., The Age Pattern of Mortality, *Journal of the Institute of Actuaries*, 1980, 107, 49-75.

Kannisto, V., *Development of Oldest-Old Mortality, 1950-1990:   Evidence from 28 Developed Countries*, 1994, Odense University.

Keyfitz, N., Choice of Function for Mortality Analysis: Effective Forecasting Depends on a Minimum Parameter Representation, *Theoretical Population Biology*, 1982, 21, 329-352.

London, D., *Graduation: The Revision of Estimates*, ACTEX Publications, Winsted, Connecticut.

Olshansky, S.J. and Carnes, B.A., Ever Since Gompertz, *Demography*, 1997, 34(1), 1-15.

Pollard, J.H., Projection of Age-specific Mortality Rates, *Population Bulletin of the United Nations*, 1987, No. 22, 55-69.

Prentice, R.L. and El Shaarawi, A., A Model for Mortality Rates and a Test of Fit for the Gompertz Force of Mortality, *Applied Statistics*, 1973, 22 (3), 301-314.

Thatcher, A.R., Some Results on the Gompertz and Heligman and Pollard Laws of Mortality, *Journal of the Institute of Actuaries*, 1990, 117, 135-149.

Thatcher, A.R., The Long-term Pattern of Adult Mortality and the Highest Attained Age, *Journal of Royal Statistical Society*, 1999, Series A, 162, 5-43.

Tuljapurkar, S. and Boe, C., Mortality Change and Forecasting:　How Much and How Little Do We Know?, *North American Actuarial Journal*, 1998, 2(4), 13-47.

Willets, R., Mortality in the Next Millennium, Paper presented in the *Staple Inn Actuarial Society*, December 1999.

Wilmoth, J., Are Mortality Rates Falling at Extremely High Ages? An Investigation Based on a Model Proposed by Coale and Kisker, *Population studies*, 1995, 49, 281-295.

Wilmoth, J., Skytthe, A., Friou, D., and Jeune, B., The Oldest Man Ever? A Case Study of Exceptional Longevity, *the Gerontologist*, 1996, 36(6), 783-788.

Wilmoth, J. and Horiuchi, S., Rectangularization Revisited:　Variability of Age at Death Within Human Populations, *Demography*, 1999, 36(4), 475-495.

Yue, C.J., *Graduation:　The Application of Statistics in Insurance.* 1997, Yeh-Yeh Bookstore, Taipei.